A New Approach for the Glottis Segmentation using Snakes

G. Andrade Miranda, N. Saenz-Lechón, V. Osma-Ruiz and J. I. Godino-Llorente
Dep. de Ingeniería de Circuitos y Sistemas, Universidad Politécnica de Madrid,
Ctra. de Valencia km. 7, 28031 Madrid, Spain
gustavo.andrade.miranda@alumnos.upm.es, {nslechon, vosma, igodino}@ics.upm.es

Keywords: Snakes, Gradient Vector Flow, Glottis Segmentation, Anisotropic Filter.

Abstract: The present work describes a new methodology for the automatic detection of the glottal space from laryngeal images based on active contour models (snakes). In order to obtain an appropriate image for the use of snakes based techniques, the proposed algorithm combines a pre-processing stage including some traditional techniques (thresholding and median filter) with more sophisticated ones such as anisotropic filtering. The value selected for the thresholding was fixed to the 85% of the maximum peak of the image histogram, and the anisotropic filter permits to distinguish two intensity levels, one corresponding to the background and the other one to the foreground (glottis). The initialization carried out is based on the magnitude obtained using the Gradient Vector Flow field, ensuring an automatic process for the selection of the initial contour. The performance of the algorithm is tested using the Pratt coefficient and compared against a manual segmentation. The results obtained suggest that this method provided results comparable with other techniques such as the proposed in (Osma-Ruiz et al., 2008).

1 INTRODUCTION

Currently, there are many works concerning the problem of the automatic detection of the glottal space as a prior step for the analysis of different phonation parameters. Roughly speaking, these works use two different approaches for the segmentation: region, and model based approaches. In the first one we find methods based on thresholded histograms and region growing ((Mehta et al., 2011), (Yan et al., 2006)).

Within the model-based approaches are the active contours, also known as snakes (Marendic et al., 2001). The snakes are thin elastic bands which are coupled appropriately to non-rigid and amorphous contours. To do that, the snake is required to be placed near the desired object (initialization), and then it is guided by external forces of the image, and once there, any additional development will not produce any change (Acton and Ray, 2009).

The snake model is controlled by two kinds of energies: external and internal. The external energy E_{ext} is generated by processing the image $I(x,y)$, producing a force that is used to drive the snake towards features of interest. Whereas the internal energy E_{int} serves to impose a piecewise smoothness constraint (Kass et al., 1988). For simplicity the $\alpha(s)$ and $\beta(s)$ parameters weights are assumed to be uniform and equal; $\alpha(s) = \alpha = \beta(s) = \beta$. The total energy of the snake is obtained by the sum of the external and internal energies:

$$E_{total} = E_{ext} + E_{int} \quad (1)$$

For the evolution process, the equation (1) must be minimized to find the minimum. In our case we use the gradient descent rule to reduce the computational load.

The rest of the work is organized as follows. Section 2 develops the methodology implemented for the glottis segmentation using snakes; pre-processing, filtering and external forces. Section 3, evaluates the results obtained using the new approach, and section 4 presents some conclusions.

2 METHODOLOGY

The proposed method allows us to individualize the glottis in laryngeal images following the scheme presented in Figure 1. The function of each block is detailed next:

2.1 Pre-processing

Before we begin the pre-processing, it is necessary to convert the original image (RGB) to a grey scale one

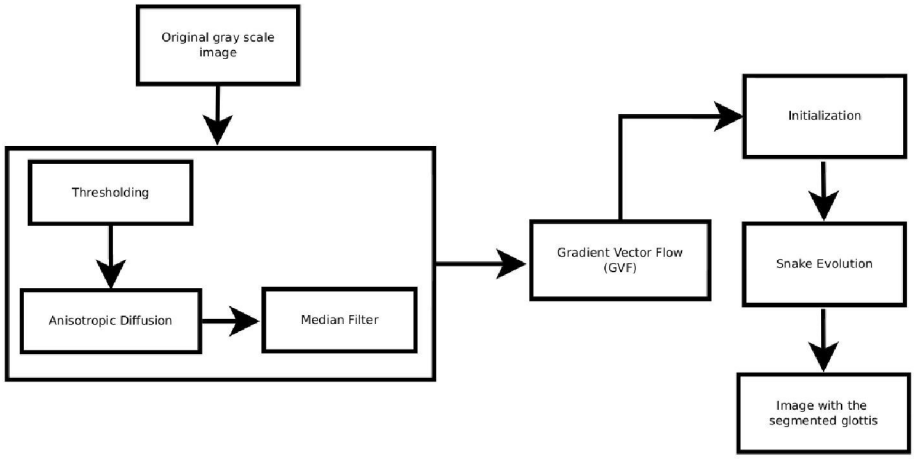

Figure 1: Diagram used for the glottis detection.

through a transformation according to the model YIQ (Russ, 2002). After such conversion, the luminance Y, is used to generate the new image in grey scale (see Figure 2). The goal of the pre-processing is to soften

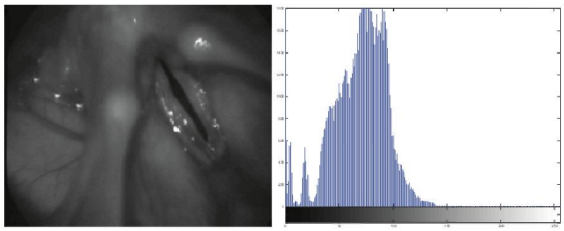

Figure 2: Laryngeal image in grey scale and its histogram.

the image and to highlight the pixels that correspond to the glottis; in this way we can avoid the snake to adjust to non desired characteristics. The pre-processing block is formed by three stages: thresholding, anisotropic diffusion and median filter.

The purpose of the thresholding stage is to reduce the contrasts obtained in the images to facilitate the job of the anisotropic filter. Based on the knowledge that the glottis is always darker than the background surrounding it, we can reassign the value of the pixels that surpass a certain amount of intensity. The selected threshold belongs to the 85% of the maximum peak at the left. This threshold was chosen to reduce as many dark pixels as possible not belonging to the glottis, and to even out the intensity in the image's background. Figure 3 shows the image obtained with its respective histogram.

Even after the thresholding step, some dark pixels continue in the surrounding of the glottis (see Figure 3), which would cause that the snake converges in wrong local minima. Therefore an extra smoothing step is necessary to even out the grey tones in the

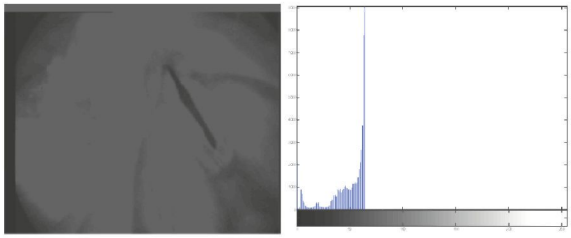

Figure 3: Thresholding and histogram of the Figure 2.

background and distinguish it from the glottis.

The objective of the anisotropic diffusion (Perona and Malik, 1990) is to soften the regions delimited by edges without affecting them, which permits us to distinguish the glottis from the background of the image. Based on (Gutiérrez-Arriola et al., 2010) we can get the desired effect in all the pixels of the obtained image during the thresholding stage. Figure 4 shows the results when we apply the anisotropic diffusion after the thresholding.

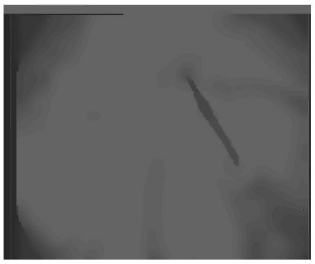

Figure 4: Anisotropic diffusion output.

A median filter replaces the grey value of a point for a median of the grey levels of its vicinity. The main goal of the median filter is to force the points isolated with intensity values that are very different from their neighbors (which in image processing is

know as salt and pepper noise) to have values closer to them. Figure 5 shows an example of an image in which a salt and pepper noise appears after the anisotropic diffusion and its respective output after the median filter.

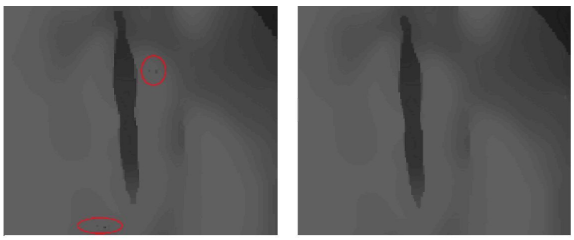

Figure 5: Anisotropic diffusion output with salt and pepper noise and median filter output.

The pre-processing reduces drastically the number and size of the local minima that don't belong to the glottis. Therefore, the problem is reduced only to seek the local minima with the biggest area to initialize the snake.

2.2 Gradient Vector Flow (GVF)

The external force GVF (Xu and Prince, 1998) is a variable of the force proposed in (Kass et al., 1988). The main idea of this force is to spread the vectors generated by (Kass et al., 1988) to its neighbors, and the neighbors at the same time to theirs. This process is done interactively along each image pixel maintaining the direction of the neighbor that generated it, and reducing its module, as it gets farther away. The GVF increases the range of the snake's movement in the image. The vector's fields that are generated through the GVF force, are used from the initialization process and evolution of the snake (see Figure 6).

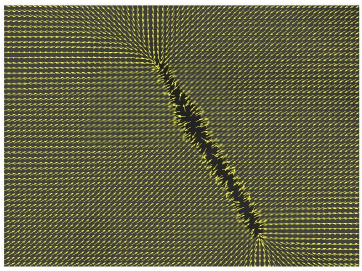

Figure 6: Vector field generated by GVF forces.

2.3 Initialization

The initialization is based on focusing on the inverse problem. In other words, what we pretend to do is to estimate the external energy from the external force GVF. The module of the vectors of the external force indicates to us how close we are to the salient feature of an image. Nevertheless, the module of the vector is not sufficient if we want to find the exact place where the glottis is located, but it is very useful when it comes to the initiation of the snake. Selecting a value of the module depends on how close from the glottis we want to initiate the snake. The experimentation let us conclude that to avoid the noise produced after the pre-processing stage, the best approach is an initialization near the glottis. The procedure is based on generating a mask with a value of 1 in those pixels that are over a threshold (0.09) and zero for the remaining ones. Thereafter, we extract the borders of the new obtained images, select the border with a bigger area, and finally we extract the coordinates corresponding to this border to place the border over the laryngeal image. Figure 7 summarizes the procedure followed.

2.4 Snake Evolution

Once we determined the initial contour, we proceed to the evolution of the snake using the lines of the GVF field. The number of iterations necessary for the snake to reach the glottis is about 50; therefore this value was used as the ending point of the iteration. The Figure 8 shows the final result of the segmentation.

3 EXPERIMENTAL RESULTS

The methodology described in the previous section has been tested with 110 images, taken from 15 videos that were recorded by the ENT service of the Gregorio Marañon Hospital in Madrid using a videostroboscopic equipment. All the images used showed the vocal folds open.

To verify the validity of our system, we did two different trials using the same database. In the first one we compared the algorithm proposed against a manual segmentation. Meanwhile, in the second one we compare with other automatic technique based on the watershed transform (Osma-Ruiz et al., 2008) against the same manual segmentation. Finally, both outcomes are compared, and the feasibility of the method proposed is discussed. The algorithm used to compare the segmentations is the Pratt algorithm. This algorithm calculates a figure of merit that measures the similarity between boundaries (Abdou and Pratt, 1979). The Pratt algorithm gives values between 0 and 1, where 1 indicates that the two edges are equal and 0 that there is no similarity at all.

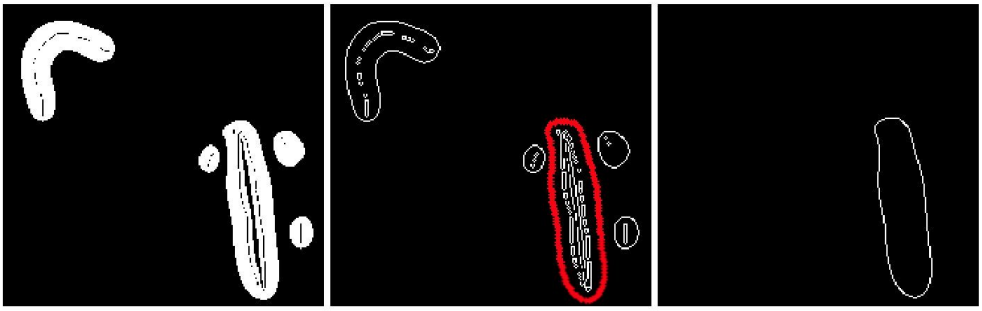

Figure 7: Initialization process.

While testing of the proposed method with the real data, we obtained 11 images with values lower than 0.5. Most of the problems that leaded to such segmentation errors were originated in the pre-processing stage. Figure 9 shows two images wrongly segmented. In the left part of each image we can observe the manual segmentation. In both images the snake only segmented a part of the glottis, this happens because the snake can not distinguish correctly between the glottis and the background.

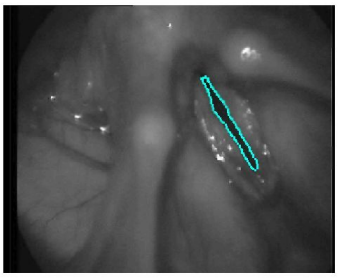

Figure 8: Final result of the segmentation.

Figure 9: Errors in the glottis detection.

The values obtained are summarized in Figure 10, through a dispersion graphic that showed the different values of the Pratt coefficient obtained for the 110 images. The images with the highest values of the coefficient are showed in Figure 11, where we can see that the difference between the manual and snake segmentations are minimal.

After testing the method based on the watershed transform, we can observe that all the Pratt coefficients are higher than 0.5 (see Figure 12). Intuitively

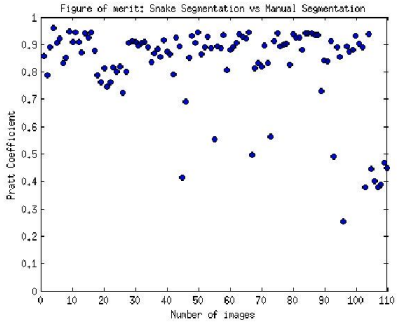

Figure 10: Summary of the Pratt coefficient obtained using method proposed.

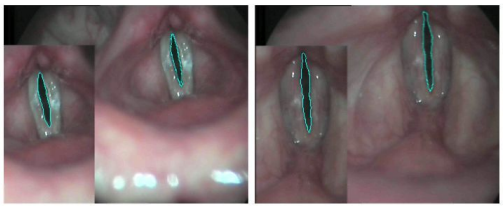

Figure 11: Images with the highest Pratt coefficient.

we would think that the aforementioned method is better than the proposed one. However this method needs to adjust the merging cost threshold in 25% of the images, whereas ours uses the same parameters for all the images. Additionally, the watershed method needs a second classification stage to detect the glottis among the rest of the objects present in the image after the merging process. Our method avoids the use of a classifier due its pre-processing step, in which the most of the objects have been deleted or reduced in size compared with the glottis. Therefore, there is no need that the system will know the shape of the glottis; identifying the object with the biggest area is enough for a successful process.

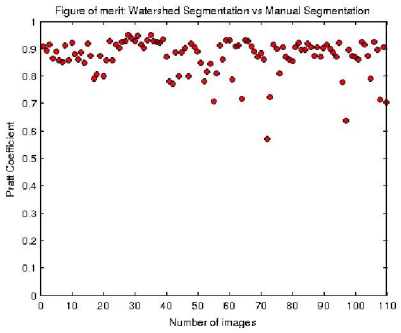

Figure 12: Summary of the Pratt coefficient obtained using method based on watershed transform (Osma-Ruiz et al., 2008).

4 CONCLUSIONS

The present work proposes an alternative to the existent methods for the glottis segmentation in laryngeal images. Despite of the poor illumination in most images, this methodology provided good results in the majority of the tested images. Only 11 images had Pratt coefficients lower than 0.5. The errors in the segmentation process are attributable to the pre-processing stage that causes the glottis to lose details and to be confused with the background. This in turn complicates the work of the snake that only segmented part of the glottis which was not affected. To resolve this inconvenience is necessary to adjust the parameters involved in the pre-processing for each image that presented errors in the segmentation. Other inconvenience presented in the method proposed, its the hard dependence of the pre-processing. All of the subsequent stages are closely related with it. Therefore a wrong setting in the parameters in the first stage could affect the remaining.

One of the most important achievements reached is the fact that we do not need to incur in heuristic criteria as the mentioned in the previous work such as: "the glottis is the darkest object in the image" or "the glottis is always centered in the image". We avoid them, based on the fact that the glottis is always surrounded by grey tones. Taking this account, we can even out the pixels that belong to the background and highlight the pixels that belong to the glottis. Lastly, but not least important is the fact that the snake can be used for tracking, whereupon the algorithm proposed could be extended to real time videos.

The solution proposed is very promising even more if we consider that can be extended to tracking of the vocal fold in real time; however this algorithm need to be tested in more different conditions in order to ensure its generalization capabilities.

ACKNOWLEDGEMENTS

This research work has been financed by the Spanish government through the project grant TEC2009-14123-C04-02.
The authors would also thank the ENT service of the Gregorio Marañon Hospital for the acquisition of the images.

REFERENCES

Abdou, I. E. and Pratt, W. K. (1979). Quantitative design and evaluation of enhancement/thresholding edge detectors. *Proceedings of The IEEE*, 67:753–763.

Acton, S. T. and Ray, N. (2009). Biomedical image analysis: Segmentation. *Synthesis Lectures on Image, Video, and Multimedia Processing*, 4(1):1–108.

Gutiérrez-Arriola, J., Osma-Ruiz, V., Godino-Llorente, J., Sáenz-Lechón, N., Fraile, R., and no, J. A.-L. (2010). Preprocesado avanzado de imágenes laríngeas para mejorar la segmentación del Área glotal. In *1er Workshop de Tecnologías Multibiométricas para la identificación de Personas*, Las Palmas de Gran Canaria, España.

Kass, M., Witkin, A., and Terzopoulos, D. (1988). Snakes: Active contour models. *INTERNATIONAL JOURNAL OF COMPUTER VISION*, 1(4):321–331.

Marendic, B., Galatsanos, N., and Bless, D. (2001). New active contour algorithm for tracking vibrating vocal folds. In *Image Processing, 2001. Proceedings. 2001 International Conference on*, volume 1, pages 397 – 400 vol.1.

Mehta, D. D., Deliyski, D. D., Quatieri, T. F., and Hillman, R. E. (2011). Automated measurement of vocal fold vibratory asymmetry from high-speed videoendoscopy recordings. *Speech, Language and Hearing Research*, 54(1):47 – 54.

Osma-Ruiz, V., Godino-Llorente, J. I., Senz-Lechn, N., and Fraile, R. (2008). Segmentation of the glottal space from laryngeal images using the watershed transform. *Computerized Medical Imaging and Graphics*, 32(3):193 – 201.

Perona, P. and Malik, J. (1990). Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.*, 12(7):629–639.

Russ, J. C. (2002). *Image Processing Handbook, Fourth Edition*. CRC Press, Inc., Boca Raton, FL, USA, 4th edition.

Xu, C. and Prince, J. L. (1998). Snakes, shapes, and gradient vector flow. *IEEE TRANSACTIONS ON IMAGE PROCESSING*, 7(3):359–369.

Yan, Y., Chen, X., and Bless, D. (2006). Automatic tracing of vocal-fold motion from high-speed digital images. *Biomedical Engineering, IEEE Transactions on*, 53(7):1394 –1400.